Hub Genes of Stroke Identified by Weighted Gene Co-expression Network Analysis

Junhong Li

First Affiliated Hospital of Guangxi University of Chinese Medicine <https://orcid.org/0000-0002-2154-8695>

Yang Zhai

GuangXi University of Chinese Medicine

Peng Wu

GuangXi University of Chinese Medicine

Yueqiang Hu

GuangXi University of Chinese Medicine

Wei Chen

GuangXi University of Chinese Medicine

Jinghui Zheng

GuangXi University of Chinese Medicine

Nong Tang (✉ nongtang0125@yeah.net)

<https://orcid.org/0000-0002-1479-2438>

Research

Keywords: Stroke, Co-expression, Network analysis, WGCNA, Hub gene

Posted Date: June 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-38160/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

BACKGROUND: Microarray-based gene expression profiling is widely used in biomedical research. Weighted gene co-expression network analysis (WGCNA) links microarray data directly to clinical traits and identifies rules for predicting pathological stage and prognosis of disease. WGCNA is useful in understanding many biological processes. Stroke is a common disease worldwide, however, molecular mechanisms of its pathogenesis are largely unknown. The aim of this study was to construct gene co-expression networks for identification of key modules and hub genes associated with stroke pathogenesis.

METHODS: Gene microarray expression profiles of stroke samples were retrieved from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) were screened by the limma package in R software. WGCNA was used to construct free-scale gene co-expression networks to explore the associations between gene sets and clinical features, and to identify key modules and hub genes. Subsequently, functional enrichment analyses were performed. Further, receiver operating characteristic (ROC) curve analysis was carried out to validate expression of hub genes and literature validation was performed as well.

RESULTS: A total of 11,747 most variant genes were used for co-expression network construction. Pink and yellow modules were significantly correlated to stroke pathogenesis. Functional enrichment analysis showed that the pink module was mainly involved in regulation of neuron regeneration, and repair of DNA damage. On the other hand, yellow module was mainly enriched in ion transport system dysfunction which was correlated with neuron death. A total of eight hub genes (PRR11, NEDD9, Notch2, RUNX1-IT1, ANP32A-IT1, ASTN2, SAMHD1 and STIM1) were identified and validated at transcriptional levels and through existing literature.

CONCLUSION: The eight hub genes (PRR11, NEDD9, Notch2, RUNX1-IT1, ANP32A-IT1, ASTN2, SAMHD1 and STIM1) identified in the study are potential biomarkers and therapeutic targets for effective diagnosis and treatment of stroke.

Background

Stroke is the second leading cause of death across the globe, after ischaemic heart disease and is predicted to be among the leading causes of death globally in the next 15 years[1,2]. It accounts for 11.3% of all deaths each year with more than 85% of stroke-related deaths occurring in low- and middle-income countries[3,4]. Previous studies report an increase in global burden of stroke has been reported despite declining mortality rates over the past two decades[5]. Current epidemiological data shows that approximately 16.9 million people are affected by stroke each year[6]. Further, epidemiological studies predict that by 2030, the number of stroke survivors will rise to 77 million[7]. Stroke is an archetypical complex disease mediated by genetic and environmental factors, therefore, the proportion of subtypes of stroke vary with age, race and ethnic origin[4,5]. Differences in subtypes among various target groups is the main therapeutic challenge in stroke management.

Microarray-based gene expression profiling is widely used in biomedical research especially in chronic non-communicable diseases including neurological diseases[8], cardiovascular disease[9], diabetes[10] and cancer[11]. In previous studies, most of microarray analysis methods focus on the comparison between normal and diseased conditions[12]. Identification of differential gene expression is a widely used analytical strategy for screening of potential biomarkers in diseased state. Several studies microarray-based gene expression have been carried out, however, single microarray dataset have high false positive rates. Therefore, integration of multiple datasets would significantly increase the reliability and sensitivity of findings and reduce false positives. Gene co-expression network analysis, the recent systems biology approach is widely used for analysis of correlation patterns among microarray analysis[13-15].

Weighted gene co-expression network analysis (WGCNA)[16,17], a tool for exploring the system-level functionality of genes, is widely used in gene expression data analysis. It defines gene co-expression by determining similarity in gene expression between expression profiles and identified groups of genes that are highly correlated across the samples [17]. WGCNA is useful in understanding various biological processes. It helps unravel the interactions between genes in different modules and hence can be used for identification of candidate biomarkers or therapeutic targets[18,19]. In addition, WGCNA links microarray data directly to clinical traits thus revealing mechanisms of drug resistance [20]. Further, WGCNA is used for identification of factors for predicting pathological stage and prognosis of disease[21].

The aim of this study was to construct co-expression modules using blood samples from stroke patients. Differentially expressed genes (DEGs) between expression profiles were identified. Further, biological function and Gene ontology (GO) enrichment analysis were carried out. Hub genes were identified and effects of biomarkers for stroke were confirmed through receiver operating characteristic (ROC) curve analysis and literature validation. Identified modules and hub genes contribute to understanding of the molecular mechanisms of stroke and are potential biomarkers for stroke gene therapy.

Materials And Methods

Microarray data

Gene expression profiles of GSE22255[22], GSE16561[23] and GSE58294[24] were obtained from Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>)[updated on June 6, 2019]. At the time of data retrieval, GSE22255 had 20 stroke samples and 20 non-stroke control samples. Further, GSE58294 contained a total of 92 blood samples including 69 stroke samples and 23 control samples. GSE16561 included 39 ischemic stroke samples and 24 healthy control samples. Raw data in this study were based on the Affymetrix platform GPL570 and Illumina platform GPL6883.

Data preprocessing and differentially expressed genes analysis

Data analysis workflow of this study is shown in sFig. 1. Analysis was performed using the R software (version 3.5.1). Raw data were processed using a Robust Multi-array Average (RMA) algorithm based on a precompiled C language function in limma package. Data preprocessing included background correcting, normalizing and calculation of expression levels[25]. Missing values in the raw data were imputed using the knn function in the impute package, and any probe absent from all CEL files was eliminated. Further, the probe-sets were annotated using annotate package. After annotation, we used the built-in match function in R to match probe-sets to their gene symbol. In cases where multiple probes matched a single gene, the probes with the highest interquartile range (IQR) were selected as described in previous studies[26]. An expression matrix with 23,495 genes was generated for subsequent analysis. The top 50% most variant genes (11,747 genes) [27] were considered to be DEGs and selected for WGCNA analysis.

WGCNA analysis

Co-expression network analysis of expression data from GSE22255 was conducted using a convenient one-step network construction method in the WGCNA package[16] to find modules associated with stroke. To ensure the reliability of expression data, samples were clustered for detection of outliers. Further, thresholding power β based on the criterion of approximate scale-free topology was selected for constructing a weighted gene network. The soft threshold calculates adjacency ranging from 0 to 1, so that the constructed network conforms to the power-law distribution and reflects the real biological network state[28]. In addition, the scale-free gene network was constructed and genes with similar patterns of expression (modules) were identified using blockwiseModules function in the WGCNA package. This function uses a dynamic tree-cutting algorithm to divide the hierarchical clustering tree into branches [29]. WGCNA approach takes the association between the two connected genes into account, and considers the topological overlap measure (TOM) which represents the overlap in shared adjacent genes. Based on the single block and block-wise module colors, we then calculated the module eigengene (ME) which represents the expression level for each module. Strength of interaction between clinical trait and ME in each module was calculated to identify the clinical significant module. Finally, gene significance (GS) in the module was defined as average p-value of each gene whereas the module significance (MS) was represented as the average GS of all genes in a given module[27].

Enrichment and modules network analysis

The ClueGO plugin in Cytoscape(version 3.7.0) (<https://cytoscape.org/>)[30] was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of clinical significant modules [31]. Significant GO terms and KEGG pathways ($P < 0.05$) were selected. Further, MCODE plugin in Cytoscape was used to construct modules network and sub-networks were extracted for further analysis.

Hub gene identification and validation

Hub genes are considered as functionally significant. The hub gene in a module was identified by module connectivity when the absolute value of the Pearson's correlation of gene-module membership (MM) was

larger than 0.9 [27,32]. In cases where the absolute value of the Pearson's correlation of gene-trait significance (TS) was more than 0.2[27], a gene was considered as highly correlated with a certain clinical trait. Furthermore, genes with both $|MM| \geq 0.9$ and $|TS| \geq 0.2$ were regarded as “real” hub genes. In the verification set of GSE16561 and GSE58294, the receiver operating characteristic (ROC) curve analysis were performed to validate the expression of hub genes in stroke and control samples. Moreover, the area under curve (AUC) was calculated using pROCR package[33]. Larger AUC value of a gene indicated that it can effectively distinguish stroke from the control samples, and the hub gene with AUC > 0.6 in the three datasets was defined as a diagnostic efficiency gene [28,34].

Results

3.1 Co-expression network construction

Samples of GSE22255 were clustered to detect outliers using hierarchical cluster analysis. One sample was removed, and 39 samples were retained (sFig. 2). Retained samples were clustered again and clinical traits were converted to color representation and visualized using heatmaps (sFig 3). In the current study, scale free topology for multiple soft thresholding powers were calculated, and the power of $\beta = 12$ (scale free $R^2 = 0.905$) was selected as an appropriate soft thresholding parameter for weighted co-expression network construction (sFig. 4). A total of 11,747 most variant genes were used for co-expression network construction. Cluster dendrogram and network heatmap of the most variant genes were constructed based on a dissimilarity measure (Figure 1).

3.2 Key modules identification

After k-means clustering, detecting, calculating and checking module eigengenes, a total of 24 modules were identified (s.Table1). Grey module was not included in any module, so the subsequent analysis was not performed on this module[28]. Lightgreen(73 genes), pink(317 genes), midnightblue(109 genes) and yellow(870 genes) modules were highly correlated with disease status (Figure 2). These modules were selected as clinically significant modules for further analysis. ME was in agreement with the expression level for each module which showed that lightgreen, pink and midnightblue modules were downregulated, while the yellow module was upregulated. Moreover, lightgreen (correlation coefficient $r = -0.41$, $P = 0.01$), pink($r = -0.33$, $P = 0.04$) and midnightblue ($r = -0.4$, $P = 0.01$) modules were negatively correlated with disease status. On the contrary, yellow($r = 0.35$, $P = 0.03$) module were positively correlated with disease status (Figure 2). In addition, correlations between GS and the two modules (pink and yellow) were 0.21($P < 0.001$) and 0.16($P < 0.001$), respectively, which indicated significant correlation with stroke (Figure 3). Therefore, pink and yellow modules were used for subsequent analysis.

3.3 GO enrichment KEGG pathway analysis of yellow and pink modules

To explore the biological processes and pathways in the pink and yellow modules, enrichment analysis was performed. GO enrichment KEGG pathway analysis were conducted using ClueGO plugin in Cytoscape (detailed information is shown in Table1). Analysis showed that biological processes of the

yellow module were mainly enriched in ion transport system dysfunction including potassium ion transport, cellular potassium ion transport and potassium ion transmembrane transport, which are implicated in neuron death. Biological processes of the pink module were mainly involved in regulation of neuron projection regeneration and repair of DNA damage such as nucleotide-excision repair, transcription-coupled nucleotide-excision repair which play an important role in prevention of cell death after stroke. In addition, pathways of the yellow module were enriched in neuroactive ligand-receptor interaction, arginine biosynthesis, and alanine, aspartate and glutamate metabolism. In the pink module, pathways were mainly enriched in nucleotide excision repair, retinol metabolism and B cell receptor signaling pathway.

3.4 Hub genes identification and validation

Based on the cut-off criteria ($|MM| > 0.9$ and $|TS| > 0.2$), 11 genes in the yellow module and 13 genes in the pink module were identified as hub genes (Table2). In GSE22255 dataset, expression levels of SAMHD1 and RUNX1-IT1 were significantly lower in stroke samples compared with control samples (sFig. 5). Moreover, ROC curve analysis indicated that 19 hub genes (all with $AUC > 0.6$) exhibited good diagnostic efficiency for control and stroke tissues (Table3). To ensure the robustness and reliability of the results, ROC analysis of the diagnostic efficiency genes was validated in two other GEO datasets, GSE16561 and GSE58294. The results showed that 8 key genes (PRR11, NEDD9, Notch2, RUNX1-IT1, ANP32A-IT1, ASTN2, SAMHD1 and STIM1) with five protein coding genes and six noncoding genes can effectively distinguish stroke from control samples (Table3). To assess the biological significance of identified hub genes, we searched literature and constructed key genes associated network using Agilent Literature Search plugin in the Cytoscape software (Figure 4). A total of 50 nodes were identified in the six networks, including 17 nodes with 26 interactions in the PRR11 network, 5 nodes with 6 interactions in the NEDD9 network, 13 nodes with 28 interactions in the Notch2 network, 4 nodes with 3 interactions in the STIM1 network, 9 nodes with 6 interactions in the SAMHD1 network and 2 nodes with 1 interactions in the ASTN2 network.

Discussion

Stroke remains a global burden to human health. Understanding molecular mechanisms implicated in stroke pathogenesis and prognosis is critical for development of precision medicine or personalized medicine. Although the pathogenesis of stroke is extremely complicated, recently, significant progress has been made in areas such as energy metabolism disorders, excitatory amino acid toxicity, penumbra depolarization and apoptosis, which are involved in stroke pathophysiological processes. However, molecular mechanisms of stroke have not been fully explored. Therefore, for effective management of stroke studies should explore the molecular mechanisms implicated in the development of stroke. In this study, we used gene expression datasets from GEO database to construct a weighted gene co-expression network. Further, we identified pathways and key genes that are potential biomarkers or therapeutic targets for stroke. In addition, whole genome data for stroke from GEO database and literature were used for validation of results.

WGCNA was performed to construct free-scale gene co-expression networks, and to explore gene co-expression modules associated with stroke. A total of 11,747 most variant genes were used for construction of co-expression network and 24 modules were identified. Pink module with 317 genes and yellow module with 870 genes had significant association with stroke. A total of 26 genes with high connectivity were screened out from the two modules. Among them, 8 hub genes were highly associated with stroke. Hub genes identified for the pink module were PRR11, NEDD9, Notch2, RUNX1-IT1 and ANP32A-IT1. On the other hand, hub genes identified for the yellow module were ASTN2, SAMHD1 and STIM1. Co-expression analysis showed that different modules were correlated with different functions. Notably, genes in the pink module are implicated in cell apoptosis and neuronal differentiation whereas genes in the yellow module are implicated in synaptic form and stroke recovery.

PRR11 (proline-rich protein 11), which is involved in cell cycle regulation, is involved in regulation of protein-protein interaction and cell signal transduction via Wnt/ β -catenin signaling pathway[35]. PRR11 consists of a binary nuclear localization signal, two proline enrichment regions and a zinc finger domain, which regulates gene transcription by binding to duplex DNA[36]. Previous studies report that PRR11 is implicated in tumor progression. However, the role and clinical application value of PRR11 in stroke is unknown[36,37]. Studies show that Notch signaling in cerebral ischemia plays a role in inflammation, oxidative stress, apoptosis, angiogenesis, synaptic plasticity and the function of blood-brain barrier[38-40]. Notably, Notch2 is upregulated with increased cell death shortly after cerebral ischemia injury in hippocampal areas [39]. Further, an increase in Notch2 levels in apoptotic cells was reported after oxygen and glucose deprivation treatment[41]. In addition, Notch2 signaling is associated with the progression of atherosclerosis[42]. Moreover, Notch2 mediates quiescence in endothelial cells, whereas inflammatory cytokines trigger increased Notch2 activity promoting apoptosis. Understanding the role of Notch2 signaling in stroke may provide insights on development of new therapeutics [43]. The brain undergoes self-repair by producing new neurons and has the ability to compensate for the loss of function after stroke[44]. NEDD9 (Neuronal precursor cell-expressed, developmentally down-regulated gene) was initially identified in mouse central nervous system[45,46]. NEDD9 is a splicing variant of Cas-L and promotes neurite outgrowth through tyrosine phosphorylation and is absent in adult brain. Interestingly, NEDD9 is upregulated for neuronal differentiation after cerebral ischemia[46]. This implies that NEDD9 is involved in recovery of neurologic function after stroke, and upregulation of NEDD9 may widen the therapeutic time window for cerebral ischemia [46,47].

SAMHD1, originally identified from a dendritic cell, is a deoxyribonucleoside triphosphate triphosphohydrolase[48-50] and its mutations were recently linked to susceptibility to stroke. Recent studies have shown that SAMHD1 gene mutations might cause genetic predispositions that interact with other risk factors, resulting in increased vulnerability to stroke[51]. Furthermore, a stroke cohort study reported that SAMHD1 gene mutations may be implicated in stroke pathogenesis in general population[52]. In addition, a functional loss of SAMHD1 protein resulting from the missense mutations c.64C>T and c.841G>A were reported to play a role in stroke pathogenesis [53]. In patients with cerebrovascular diseases, lack of SAMHD1 protein expression was associated with decreased expression of IFNB1 and increased expression of IL8[54].

ASTN2 (astrotactin-2) is a conserved perforin-like membrane protein expressed in the developing and adult brain and is primarily involved in neuronal development[55,56]. ASTN2 modulates a number of protein complexes in neurons that impact synaptic form and regulate synaptic adhesion activity[57]. ASTN2 binds to various interacting proteins, like ROCK2 and SLC12a5, which are implicated in vesicle trafficking and synaptic function. Ca^{2+} entry is important for platelet activation and stroke. Glutamate-induced dysregulation of intracellular Ca^{2+} homeostasis is a key mechanism in stroke pathogenesis[58]. STIM1 (stromal interaction molecule 1), a Ca^{2+} sensor localized in endoplasmic reticulum, is involved in regulation of store-operated calcium entry[59]. Defective Ca^{2+} entry mechanism in mutant platelets activate STIM1, resulting in impaired platelet aggregate formation[60]. This process offers protection to mice from ischemic stroke[61]. Dong M et al reported that STIM1/Orai1 expression was associated with mortality and recurrence in ischemic stroke patients and was an independent predictor of the 3-month stroke recovery [62].

Intriguingly, three hub genes were identified in this study. Recent studies have shown that long non-coding RNA is considered a key regulator of pathogenesis of stroke[63]. Two key noncoding RNAs (RUNX1-IT1 and ANP32A-IT1) were identified as effective diagnostic genes in the present study. LncRNA RUNX1-IT1, and ANP32A-IT1 are the intronic transcript 1 from their respective genes[64,65]. RUNX1 IT1 plays a tumour suppressive role and plays a protective role in colorectal cancer by regulating cell proliferation, migration, and apoptosis[65]. Although the roles of RUNX1-IT1 in different diseases have been explored [66], its biological roles in stroke remains unknown. ANP32A is implicated in neurons differentiation, brain development, and neuritogenesis. Modulating ANP32A signaling could help manage oxidative stress in brain[67] and restore cognitive function[68] with therapeutic implications for neurological disease. To our best knowledge, no study has reported the role of RUNX1-IT1 and ANP32A-IT1 in stroke.

Neuronal death in cerebral ischemic area of stroke is accompanied by unregulated genes expression. Therefore, the hub genes of the pink module can regulate cellular signal transduction involved in inflammation, oxidative stress, apoptosis, angiogenesis, and synaptic plasticity. On the other hand, hub genes of the yellow module may play a role in synaptic form, antiplatelet aggregate formation and short-term prognosis after stroke. Therefore, we speculated that pink module and yellow module are key modules in neuronal death, new synapse formation, and functional recovery of stroke. Moreover, non-coding genes subset in the hub genes were identified in the present study. Non-coding RNAs play a key regulatory role in pathogenesis of stroke, therefore, further studies involving non-coding RNAs populations and exploring the mechanism underlying their role in stroke should be carried out.

Co-expression analyses revealed mRNA regulation expression network in stroke. In the present study, we used WGCNA to construct gene co-expression networks and to explore the relationships between modules and clinical traits. Due to the high degree of consistency in the expression relationships of the module, genes of the same module might share a common biological role. Our results provide valuable information for basic and clinical research of stroke. Although there are similar limitations with most other data mining approaches[69], we adopted the indicated methods to reduce possible bias. In order to

increase the credibility of WGCNA results, we used matched stroke and normal samples for analysis, and external data from GEO database to validate the results.

Conclusion

The hub genes PRR11, NEDD9, Notch2, RUNX1-IT1, ANP32A-IT1, ASTN2, SAMHD1 and STIM1 were found to be significantly correlated with stroke and were verified in other two datasets. These hub genes can be used as potential diagnostic and prognostic biomarkers of stroke, although further research should be carried out. Moreover, pink and yellow modules were considered as critical modules in neuronal death, new synapse formation, and functional recovery of stroke.

Declarations

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Yes.

Availability of data and materials:

Not applicable.

Competing interests:

None.

Funding:

This study was funded by Guangxi Science and Technology Project(grant number: 2019JJB140151), Guangxi Key Laboratory of Chinese Medicine Foundation Research Project(grant number: K201255808 and K201137906), Innovation Project of Guangxi Graduate Education(grant number: YCBZ2020067), Guangxi first-class discipline construction project(grant number: 2019XK025 and 2019XK024).

Authors' contributions:

Junhong Li, Yang Zhai, Yueqiang Hu, Wei Chen, and Nong Tang involved in study concept and design. Junhong Li and Yang Zhai involved in drafting of the manuscript. All authors involved in data acquisition, analysis, or interpretation and critical review of the manuscript.

Acknowledgements:

Not applicable.

References

- [1] W.H. Organization, The Top 10 Causes of Death, <http://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death>, 2018 (accessed May 24 2018).
- [2] C.M. Stinear, Prediction of motor recovery after stroke: advances in biomarkers, *Lancet Neurol* 16 (2017) 826-836. 10.1016/S1474-4422(17)30283-1.
- [3] E.B. Limbole, J. Magne, P. Lacroix, Stroke characterization in Sun Saharan Africa: Congolese population, *Int J Cardiol* 240 (2017) 392-397. 10.1016/j.ijcard.2017.04.063.
- [4] M. Writing Group, D. Mozaffarian, E.J. Benjamin, A.S. Go, D.K. Arnett, M.J. Blaha, M. Cushman, S.R. Das, S. de Ferranti, J.P. Despres, H.J. Fullerton, V.J. Howard, M.D. Huffman, C.R. Isasi, M.C. Jimenez, S.E. Judd, B.M. Kissela, J.H. Lichtman, L.D. Lisabeth, S. Liu, R.H. Mackey, D.J. Magid, D.K. McGuire, E.R. Mohler, 3rd, C.S. Moy, P. Muntner, M.E. Mussolino, K. Nasir, R.W. Neumar, G. Nichol, L. Palaniappan, D.K. Pandey, M.J. Reeves, C.J. Rodriguez, W. Rosamond, P.D. Sorlie, J. Stein, A. Towfighi, T.N. Turan, S.S. Virani, D. Woo, R.W. Yeh, M.B. Turner, C. American Heart Association Statistics, S. Stroke Statistics, Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association, *Circulation* 133 (2016) e38-360. 10.1161/CIR.0000000000000350.
- [5] G.J. Hankey, Stroke, *Lancet* 389 (2017) 641-654. 10.1016/S0140-6736(16)30962-X.
- [6] V.L. Feigin, M.H. Forouzanfar, R. Krishnamurthi, G.A. Mensah, M. Connor, D.A. Bennett, A.E. Moran, R.L. Sacco, L. Anderson, T. Truelsen, M. O'Donnell, N. Venketasubramanian, S. Barker-Collo, C.M. Lawes, W. Wang, Y. Shinohara, E. Witt, M. Ezzati, M. Naghavi, C. Murray, I. Global Burden of Diseases, S. Risk Factors, G.B.D.S.E.G. the, Global and regional burden of stroke during 1990-2010: findings from the Global Burden of Disease Study 2010, *Lancet* 383 (2014) 245-254.
- [7] Y. Bejot, B. Daubail, M. Giroud, Epidemiology of stroke and transient ischemic attacks: Current knowledge and perspectives, *Rev Neurol (Paris)* 172 (2016) 59-68. 10.1016/j.neurol.2015.07.013.
- [8] H. Nakaoka, A. Tajima, T. Yoneyama, K. Hosomichi, H. Kasuya, T. Mizutani, I. Inoue, Gene expression profiling reveals distinct molecular signatures associated with the rupture of intracranial aneurysm, *Stroke* 45 (2014) 2239-2245. 10.1161/STROKEAHA.114.005851.
- [9] M. Steenman, G. Lamirault, N. Le Meur, J.J. Leger, Gene expression profiling in human cardiovascular disease, *Clin Chem Lab Med* 43 (2005) 696-701. 10.1515/CCLM.2005.118.

- [10] J. Chen, Y. Meng, J. Zhou, M. Zhuo, F. Ling, Y. Zhang, H. Du, X. Wang, Identifying candidate genes for Type 2 Diabetes Mellitus and obesity through gene expression profiling in multiple tissues or cells, *J Diabetes Res* 2013 (2013) 970435. 10.1155/2013/970435.
- [11] F. Rojo, L. Domingo, M. Sala, S. Zazo, C. Chamizo, S. Menendez, O. Arpi, J.M. Corominas, R. Bragado, S. Servitja, I. Tusquets, L. Nonell, F. Macia, J. Martinez, A. Rovira, J. Albanell, X. Castells, Gene expression profiling in true interval breast cancer reveals overactivation of the mTOR signaling pathway, *Cancer Epidemiol Biomarkers Prev* 23 (2014) 288-299. 10.1158/1055-9965.EPI-13-0761.
- [12] L. Yip, R. Fuhlbrigge, M.A. Atkinson, C.G. Fathman, Impact of blood collection and processing on peripheral blood gene expression profiling in type 1 diabetes, *BMC Genomics* 18 (2017) 636. 10.1186/s12864-017-3949-2.
- [13] J.W. Liang, Z.Y. Fang, Y. Huang, Z.Y. Liuyang, X.L. Zhang, J.L. Wang, H. Wei, J.Z. Wang, X.C. Wang, J. Zeng, R. Liu, Application of Weighted Gene Co-Expression Network Analysis to Explore the Key Genes in Alzheimer's Disease, *J Alzheimers Dis* 65 (2018) 1353-1364. 10.3233/JAD-180400.
- [14] H. Shi, L. Zhang, Y. Qu, L. Hou, L. Wang, M. Zheng, Prognostic genes of breast cancer revealed by gene co-expression network analysis, *Oncol Lett* 14 (2017) 4535-4542. 10.3892/ol.2017.6779.
- [15] J.A. Chen, Gene Co-Expression Network Analysis Implicates microRNA Processing in Parkinson's Disease Pathogenesis, *Neurodegener Dis* 18 (2018) 191-199. 10.1159/000490427.
- [16] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinformatics* 9 (2008) 559. 10.1186/1471-2105-9-559.
- [17] B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, *Stat Appl Genet Mol Biol* 4 (2005) Article17. 10.2202/1544-6115.1128.
- [18] M. Giulietti, G. Occhipinti, A. Righetti, M. Bracci, A. Conti, A. Ruzzo, E. Cerigioni, T. Cacciamani, G. Principato, F. Piva, Emerging Biomarkers in Bladder Cancer Identified by Network Analysis of Transcriptomic Data, *Front Oncol* 8 (2018) 450. 10.3389/fonc.2018.00450.
- [19] R.L. Costa, M. Boroni, M.A. Soares, Distinct co-expression networks using multi-omic data reveal novel interventional targets in HPV-positive and negative head-and-neck squamous cell cancer, *Sci Rep* 8 (2018) 15254. 10.1038/s41598-018-33498-5.

- [20] Y. Chen, F. Bi, Y. An, Q. Yang, Coexpression network analysis identified Kruppel-like factor 6 (KLF6) association with chemosensitivity in ovarian cancer, *J Cell Biochem* (2018). 10.1002/jcb.27567.
- [21] L. Chen, L. Yuan, K. Qian, G. Qian, Y. Zhu, C.L. Wu, H.C. Dan, Y. Xiao, X. Wang, Identification of Biomarkers Associated With Pathological Stage and Prognosis of Clear Cell Renal Cell Carcinoma by Co-expression Network Analysis, *Front Physiol* 9 (2018) 399. 10.3389/fphys.2018.00399.
- [22] T. Krug, J.P. Gabriel, R. Taipa, B.V. Fonseca, S. Domingues-Montanari, I. Fernandez-Cadenas, H. Manso, L.O. Gouveia, J. Sobral, I. Albergaria, G. Gaspar, J. Jimenez-Conde, R. Rabionet, J.M. Ferro, J. Montaner, A.M. Vicente, M.R. Silva, I. Matos, G. Lopes, S.A. Oliveira, TTC7B emerges as a novel risk factor for ischemic stroke through the convergence of several genome-wide approaches, *J Cereb Blood Flow Metab* 32 (2012) 1061-1072. 10.1038/jcbfm.2012.24.
- [23] T.L. Barr, Y. Conley, J. Ding, A. Dillman, S. Warach, A. Singleton, M. Matarin, Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling, *Neurology* 75 (2010) 1009-1014. 10.1212/WNL.0b013e3181f2b37f.
- [24] B. Stamova, G.C. Jickling, B.P. Ander, X. Zhan, D. Liu, R. Turner, C. Ho, J.C. Khoury, C. Bushnell, A. Pancioli, E.C. Jauch, J.P. Broderick, F.R. Sharp, Gene expression in peripheral immune cells following cardioembolic stroke is sexually dimorphic, *PLoS One* 9 (2014) e102550. 10.1371/journal.pone.0102550.
- [25] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res* 43 (2015) e47. 10.1093/nar/gkv007.
- [26] X. Zhang, H. Feng, Z. Li, D. Li, S. Liu, H. Huang, M. Li, Application of weighted gene co-expression network analysis to identify key modules and hub genes in oral squamous cell carcinoma tumorigenesis, *Onco Targets Ther* 11 (2018) 6001-6021. 10.2147/OTT.S171791.
- [27] J. Tang, D. Kong, Q. Cui, K. Wang, D. Zhang, Y. Gong, G. Wu, Prognostic Genes of Breast Cancer Identified by Gene Co-expression Network Analysis, *Front Oncol* 8 (2018) 374. 10.3389/fonc.2018.00374.
- [28] Q. Yang, R. Wang, B. Wei, C. Peng, L. Wang, G. Hu, D. Kong, C. Du, Candidate Biomarkers and Molecular Mechanism Investigation for Glioblastoma Multiforme Utilizing WGCNA, *Biomed Res Int* 2018 (2018) 4246703. 10.1155/2018/4246703.
- [29] P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R, *Bioinformatics* 24 (2008) 719-720.

10.1093/bioinformatics/btm563.

[30] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res* 13 (2003) 2498-2504. 10.1101/gr.1239303.

[31] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS* 16 (2012) 284-287. 10.1089/omi.2011.0118.

[32] L. Yuan, L. Chen, K. Qian, G. Qian, C.L. Wu, X. Wang, Y. Xiao, Co-expression network analysis identified six hub genes in association with progression and prognosis in human clear cell renal cell carcinoma (ccRCC), *Genom Data* 14 (2017) 132-140. 10.1016/j.gdata.2017.10.006.

[33] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Muller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinformatics* 12 (2011) 77. 10.1186/1471-2105-12-77.

[34] F. Long, J.H. Su, B. Liang, L.L. Su, S.J. Jiang, Identification of Gene Biomarkers for Distinguishing Small-Cell Lung Cancer from Non-Small-Cell Lung Cancer Using a Network-Based Approach, *Biomed Res Int* 2015 (2015) 685303. 10.1155/2015/685303.

[35] P. Yang, Z. Qiu, Y. Jiang, L. Dong, W. Yang, C. Gu, G. Li, Y. Zhu, Silencing of cZNF292 circular RNA suppresses human glioma tube formation via the Wnt/beta-catenin signaling pathway, *Oncotarget* 7 (2016) 63449-63455. 10.18632/oncotarget.11523.

[36] W. Qiao, H. Wang, X. Zhang, K. Luo, Proline-rich protein 11 silencing inhibits hepatocellular carcinoma growth and epithelial-mesenchymal transition through beta-catenin signaling, *Gene* 681 (2019) 7-14. 10.1016/j.gene.2018.09.036.

[37] J. Zhu, H. Hu, J. Wang, Y. Yang, P. Yi, PRR11 Overexpression Facilitates Ovarian Carcinoma Cell Proliferation, Migration, and Invasion Through Activation of the PI3K/AKT/beta-Catenin Pathway, *Cell Physiol Biochem* 49 (2018) 696-705. 10.1159/000493034.

[38] Z. Cai, B. Zhao, Y. Deng, S. Shangguan, F. Zhou, W. Zhou, X. Li, Y. Li, G. Chen, Notch signaling in cerebrovascular diseases (Review), *Molecular medicine reports* 14 (2016) 2883-2898. 10.3892/mmr.2016.5641.

[39] T.V. Arumugam, S.H. Baik, P. Balaganapathy, C.G. Sobey, M.P. Mattson, D.G. Jo, Notch signaling and neuronal death in stroke, *Prog Neurobiol* 165-167 (2018) 103-116. 10.1016/j.pneurobio.2018.03.002.

- [40] S.M. Peterson, J.E. Turner, A. Harrington, J. Davis-Knowlton, V. Lindner, T. Gridley, C.P.H. Vary, L. Liaw, Notch2 and Proteomic Signatures in Mouse Neointimal Lesion Formation, *Arterioscler Thromb Vasc Biol* 38 (2018) 1576-1593. 10.1161/ATVBAHA.118.311092.
- [41] L. Alberi, Z. Chi, S.D. Kadam, J.D. Mulholland, V.L. Dawson, N. Gaiano, A.M. Comi, Neonatal stroke in mice causes long-term changes in neuronal Notch-2 expression that may contribute to prolonged injury, *Stroke* 41 (2010) S64-71. 10.1161/STROKEAHA.110.595298.
- [42] J. Davis-Knowlton, J.E. Turner, A. Turner, S. Damian-Loring, N. Hagler, T. Henderson, I.F. Emery, K. Bond, C.W. Duarte, C.P.H. Vary, J. Eldrup-Jorgensen, L. Liaw, Characterization of smooth muscle cells from human atherosclerotic lesions and their responses to Notch signaling, *Lab Invest* 99 (2019) 290-304. 10.1038/s41374-018-0072-1.
- [43] J.L. Ables, J.J. Breunig, A.J. Eisch, P. Rakic, Not(ch) just development: Notch signalling in the adult brain, *Nat Rev Neurosci* 12 (2011) 269-283. 10.1038/nrn3024.
- [44] Y. Sun, X. Sun, H. Qu, S. Zhao, T. Xiao, C. Zhao, Neuroplasticity and behavioral effects of fluoxetine after experimental stroke, *Restorative neurology and neuroscience* 35 (2017) 457-468. 10.3233/RNN-170725.
- [45] O. Lindvall, Z. Kokaia, Neurogenesis following Stroke Affecting the Adult Brain, *Cold Spring Harb Perspect Biol* 7 (2015). 10.1101/cshperspect.a019034.
- [46] T. Sasaki, S. Iwata, H.J. Okano, Y. Urasaki, J. Hamada, H. Tanaka, N.H. Dang, H. Okano, C. Morimoto, Nedd9 protein, a Cas-L homologue, is upregulated after transient global ischemia in rats: possible involvement of Nedd9 in the differentiation of neurons after ischemia, *Stroke* 36 (2005) 2457-2462. 10.1161/01.STR.0000185672.10390.30.
- [47] E. Shagisultanova, A.V. Gaponova, R. Gabbasov, E. Nicolas, E.A. Golemis, Preclinical and clinical studies of the NEDD9 scaffold protein in cancer and other diseases, *Gene* 567 (2015) 1-11. 10.1016/j.gene.2015.04.086.
- [48] D.C. Goldstone, V. Ennis-Adeniran, J.J. Hedden, H.C. Groom, G.I. Rice, E. Christodoulou, P.A. Walker, G. Kelly, L.F. Haire, M.W. Yap, L.P. de Carvalho, J.P. Stoye, Y.J. Crow, I.A. Taylor, M. Webb, HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase, *Nature* 480 (2011) 379-382. 10.1038/nature10623.
- [49] K. Hrecka, C. Hao, M. Gierszewska, S.K. Swanson, M. Kesik-Brodacka, S. Srivastava, L. Florens, M.P. Washburn, J. Skowronski, Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein, *Nature* 474 (2011) 658-661. 10.1038/nature10195.

- [50] N. Li, W. Zhang, X. Cao, Identification of human homologue of mouse IFN-gamma induced protein from human dendritic cells, *Immunol Lett* 74 (2000) 221-224.
- [51] X. Ji, Y. Wu, J. Yan, J. Mehrens, H. Yang, M. DeLucia, C. Hao, A.M. Gronenborn, J. Skowronski, J. Ahn, Y. Xiong, Mechanism of allosteric activation of SAMHD1 by dGTP, *Nat Struct Mol Biol* 20 (2013) 1304-1309. 10.1038/nsmb.2692.
- [52] W. Li, B. Xin, J. Yan, Y. Wu, B. Hu, L. Liu, Y. Wang, J. Ahn, J. Skowronski, Z. Zhang, Y. Wang, H. Wang, SAMHD1 Gene Mutations Are Associated with Cerebral Large-Artery Atherosclerosis, *Biomed Res Int* 2015 (2015) 739586. 10.1155/2015/739586.
- [53] B. Xin, S. Jones, E.G. Puffenberger, C. Hinze, A. Bright, H. Tan, A. Zhou, G. Wu, J. Vargus-Adams, D. Agamanolis, H. Wang, Homozygous mutation in SAMHD1 gene causes cerebral vasculopathy and early onset stroke, *Proc Natl Acad Sci U S A* 108 (2011) 5372-5377. 10.1073/pnas.1014265108.
- [54] H. Thiele, M. du Moulin, K. Barczyk, C. George, W. Schwindt, G. Nurnberg, M. Frosch, G. Kurlemann, J. Roth, P. Nurnberg, F. Rutsch, Cerebral arterial stenoses and stroke: novel features of Aicardi-Goutieres syndrome caused by the Arg164X mutation in SAMHD1 are associated with altered cytokine expression, *Hum Mutat* 31 (2010) E1836-1850. 10.1002/humu.21357.
- [55] P.M. Wilson, R.H. Fryer, Y. Fang, M.E. Hatten, Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration, *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30 (2010) 8529-8540. 10.1523/JNEUROSCI.0032-10.2010.
- [56] T. Ni, K. Harlos, R. Gilbert, Structure of astrotactin-2: a conserved vertebrate-specific and perforin-like membrane protein involved in neuronal development, *Open Biol* 6 (2016). 10.1098/rsob.160053.
- [57] H. Behesti, T.R. Fore, P. Wu, Z. Horn, M. Leppert, C. Hull, M.E. Hatten, ASTN2 modulates synaptic strength by trafficking and degradation of surface proteins, *Proc Natl Acad Sci U S A* 115 (2018) E9717-E9726. 10.1073/pnas.1809382115.
- [58] Q. Song, W.L. Gou, Y.L. Zou, FAM3A Protects Against Glutamate-Induced Toxicity by Preserving Calcium Homeostasis in Differentiated PC12 Cells, *Cell Physiol Biochem* 44 (2017) 2029-2041. 10.1159/000485943.
- [59] W. Bergmeier, A.K. Chauhan, D.D. Wagner, Glycoprotein Ibalpha and von Willebrand factor in primary platelet adhesion and thrombus formation: lessons from mutant mice, *Thromb Haemost* 99 (2008) 264-270. 10.1160/TH07-10-0638.

- [60] H. Ohara, T. Nabika, A nonsense mutation of Stim1 identified in stroke-prone spontaneously hypertensive rats decreased the store-operated calcium entry in astrocytes, *Biochem Biophys Res Commun* 476 (2016) 406-411. 10.1016/j.bbrc.2016.05.134.
- [61] S.K. Gotru, W. Chen, P. Kraft, I.C. Becker, K. Wolf, S. Stritt, S. Zierler, H.M. Hermanns, D. Rao, A.L. Perraud, C. Schmitz, R.P. Zahedi, P.J. Noy, M.G. Tomlinson, T. Dandekar, M. Matsushita, V. Chubakov, T. Gudermann, G. Stoll, B. Nieswandt, A. Braun, TRPM7 Kinase Controls Calcium Responses in Arterial Thrombosis and Stroke in Mice, *Arterioscler Thromb Vasc Biol* 38 (2018) 344-352. 10.1161/ATVBAHA.117.310391.
- [62] M. Dong, N. Zheng, L.J. Ren, H. Zhou, J. Liu, Increased expression of STIM1/Orai1 in platelets of stroke patients predictive of poor outcomes, *Eur J Neurol* 24 (2017) 912-919. 10.1111/ene.13304.
- [63] L. Yang, H. Wang, Q. Shen, L. Feng, H. Jin, Long non-coding RNAs involved in autophagy regulation, *Cell death & disease* 8 (2017) e3073. 10.1038/cddis.2017.464.
- [64] J. Ramsey, K. Butnor, Z. Peng, T. Leclair, J. van der Velden, G. Stein, J. Lian, C.M. Kinsey, Loss of RUNX1 is associated with aggressive lung adenocarcinomas, *J Cell Physiol* 233 (2018) 3487-3497. 10.1002/jcp.26201.
- [65] J. Shi, X. Zhong, Y. Song, Z. Wu, P. Gao, J. Zhao, J. Sun, J. Wang, J. Liu, Z. Wang, Long non-coding RNA RUNX1-IT1 plays a tumour-suppressive role in colorectal cancer by inhibiting cell proliferation and migration, *Cell Biochem Funct* 37 (2019) 11-20. 10.1002/cbf.3368.
- [66] K. Yang, Y. Hou, A. Li, Z. Li, W. Wang, H. Xie, Z. Rong, G. Lou, K. Li, Identification of a six-lncRNA signature associated with recurrence of ovarian cancer, *Sci Rep* 7 (2017) 752. 10.1038/s41598-017-00763-y.
- [67] F.M.F. Cornelis, S. Monteagudo, L.K.A. Guns, W. den Hollander, R. Nelissen, L. Storms, T. Peeters, I. Jonkers, I. Meulenbelt, R.J. Lories, ANP32A regulates ATM expression and prevents oxidative stress in cartilage, brain, and bone, *Sci Transl Med* 10 (2018). 10.1126/scitranslmed.aar8426.
- [68] G.S. Chai, Q. Feng, R.H. Ma, X.H. Qian, D.J. Luo, Z.H. Wang, Y. Hu, D.S. Sun, J.F. Zhang, X. Li, X.G. Li, D. Ke, J.Z. Wang, X.F. Yang, G.P. Liu, Inhibition of Histone Acetylation by ANP32A Induces Memory Deficits, *J Alzheimers Dis* 63 (2018) 1537-1546. 10.3233/JAD-180090.
- [69] Q. Wan, J. Tang, Y. Han, D. Wang, Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma, *Exp Eye Res* 166 (2018) 13-20. 10.1016/j.exer.2017.10.007.

Tables

Table1

GO enrichment analysis of yellow and pink module (top 4 in each module were listed)

Module	GO ID	GO Term	P value	Associated Genes
yellow	GO:0030001	metal ion transport	0.0013	ABCC9, AGXT, ANK2, APLNR, AQP1, ASIC1, ATP12A, ATP13A5, ATP7B, CACNA1H, CACNG6, CASQ2, CMPK1, CYP27B1, DMD, DPP6, EPPIN, F2RL3, FGF2, GAL, GRIN2B, GRM6, KCNA5, KCNG4, KCNJ1, KCNJ3, KCNV1, NCOA6, NETO1, NMUR2, NOS3, PANX1, PARP3, PKD2, PPOX, SLC12A5, SLC13A3, SLC17A2, SLC23A1, SLC39A9, SLC41A2, SLC4A8, SLC5A7, SLC9A8, SNAP91, TRPA1, UCN
yellow	GO:0098660	inorganic ion transmembrane transport	0.0169	ABCC9, ANK2, APLNR, AQP1, ASIC1, ATP12A, ATP13A5, ATP7B, CACNA1H, CACNG6, CASQ2, CFTR, CLCC1, CLDN4, CMPK1, DMD, DPP6, F2RL3, FGF2, G6PC, GABRG1, GAL, GRIN2B, KCNA5, KCNG4, KCNJ1, KCNJ3, KCNV1, NCOA6, NETO1, PANX1, PARP3, PKD2, SLC12A5, SLC17A2, SLC23A1, SLC41A2, SLC9A8, SNAP91, TRPA1
yellow	GO:0009617	response to bacterium	0.0179	ABCC11, ADIPOQ, ALAD, ARHGEF28, ASS1, C10orf99, CHST4, CMPK1, CPS1, CRP, CYP1A2, CYP27B1, DCN, DEFB106A, EPPIN, FMO1, GKN2, HERC6, KYAT1, LIAS, LYPD8, MBL2, MRO, NLRP1, NOS3, PLSCR4, PNLIPRP2, PRL, RARA, RNASEH2A, S100A7A, SPAG11A, TINAGL1, TRAPPC2, UGT1A1
yellow	GO:0010035	response to inorganic substance	0.0080	ABCC11, ALAD, AQP1, ASS1, ATP7B, BAD, CACNA1H, CASQ2, CPO, CPS1, CYP1A2, DMD, EGFR, F2RL3, FXN, HBB, KCNA5, NCOA6, NOS3, PKD2, PON1, RASGRP2, RNASEH2A, S100A16, SLC17A2, SNAP91, TAT, TFAP2A, TINAGL1, TNNT2, TRAF2, TRPA1, TSHB
pink	GO:0000186	activation of MAPKK activity	0.0001	JAK2, KIDINS220, MAP4K4, RAF1, TAOK1, ZHX2
pink	GO:0006289	nucleotide-excision repair	0.0021	CUL4A, GTF2H3, POLK, POLR2J, RPA4, TCEA1
pink	GO:0031102	neuron projection regeneration	0.0004	ATF7IP, JAK2, MAP4K4, RTN4, TEP1
pink	GO:0006283	transcription-coupled nucleotide-excision repair	0.0023	CUL4A, GTF2H3, POLK, POLR2J, TCEA1

yellow	KEGG:04080	Neuroactive ligand-receptor interaction	0.0119	APLNR, CGA, F2RL3, GABRG1, GRIN2B, GRM6, HTR1E, LEPR, LHCGR, NMUR2, PRL, TAAR2, TAAR9, TRHR, TSHB
yellow	KEGG:00250	Alanine, aspartate and glutamate metabolism	0.0142	AGXT, AGXT2, ASS1, CPS1
yellow	KEGG:00220	Arginine biosynthesis	0.0181	ASS1, CPS1, NOS3
yellow	KEGG:05410	Hypertrophic cardiomyopathy (HCM)	0.0246	CACNG6, DMD, ITGA10, ITGB5, TNNT2, TPM4
pink	KEGG:03440	Homologous recombination	0.0045	ATM, RPA4, XRCC2
pink	KEGG:03420	Nucleotide excision repair	0.0066	CUL4A, GTF2H3, RPA4
pink	KEGG:00830	Retinol metabolism	0.0174	ADH4, CYP3A4, RDH11
pink	KEGG:04662	B cell receptor signaling pathway	0.0203	DAPP1, INPP5D, RAF1
<i>GO</i> gene-ontology, <i>BP</i> biological process, <i>KEGG</i> Kyoto Encyclopedia of Genes and Genomes				

Table2
Hub genes in the yellow and pink modules

Module Color	Gene Symbol	MM	p-value of MM
yellow	SAMHD1	0.9212	0.0000
yellow	POLK	0.9171	0.0000
yellow	LINC01192	0.9098	0.0000
yellow	NEGR1-IT1	0.9315	0.0000
yellow	LOC101927815	0.9120	0.0000
yellow	INO80C	0.9025	0.0000
yellow	CACNG6	0.9135	0.0000
yellow	ASTN2	0.9332	0.0000
yellow	ANKRD20A1	0.9357	0.0000
yellow	CCDC169	0.9400	0.0000
yellow	ZNF33A	0.9024	0.0000
pink	FAM49B	0.9100	0.0000
pink	N4BP2L2	0.9430	0.0000
pink	SUPT20H	0.9212	0.0000
pink	MACF1	0.9371	0.0000
pink	C4orf29	0.9223	0.0000
pink	NEDD9	0.9256	0.0000
pink	Notch2	0.9102	0.0000
pink	KDM4C	0.9236	0.0000
pink	PRR11	0.9116	0.0000
pink	VTI1A	0.9256	0.0000
pink	RUNX1-IT1	0.9230	0.0000
pink	FOXP1-IT1	0.9100	0.0000
pink	ANP32A-IT1	0.9024	0.0000
<i>MM</i> Pearson's correlation of gene-module membership			

Table3
AUC of hub genes in three datasets

Hub genes	Module	Diagnostic efficiency	GSE22255	GSE16561	GSE58294
FAM49B	pink	NO	0.698	0.628	0.473
VTI1A	pink	NO	0.608	-	0.544
RUNX1-IT1	pink	YES	0.743	-	0.969
PRR11	pink	YES	0.645	0.634	0.885
ANP32A-IT1	pink	YES	0.665	-	0.743
KDM4C	pink	YES	0.723	0.553	0.831
N4BP2L2	pink	YES	0.783	0.685	0.576
SUPT20H	pink	YES	0.718	0.597	0.677
MACF1	pink	YES	0.618	0.449	0.883
FOXP1-IT1	pink	YES	0.648	0.534	0.71
NEDD9	pink	YES	0.628	0.612	0.645
NOTCH2	pink	YES	0.651	-	0.601
C4orf29	pink	YES	0.705	-	0.543
CACNG6	yellow	NO	0.652	0.482	0.649
INO80C	yellow	NO	0.665	0.535	0.507
LOC101927815	yellow	NO	0.568	-	0.535
ASTN2	yellow	YES	0.635	0.618	0.955
SAMHD1	yellow	YES	0.627	-	0.842
POLK	yellow	YES	0.652	-	0.805
STIM1	yellow	YES	0.655	-	0.675
CCDC169	yellow	YES	0.583	-	0.7
NEGR1-IT1	yellow	YES	0.6	-	0.678
ANKRD20A1	yellow	YES	0.52	0.592	0.791
ZNF33A	yellow	YES	0.547	0.656	0.698

Figure 1

Cluster dendrogram and network heatmap plot. A. Dendograms produced by dissimilarity based on topological overlaps with assigned module colors. The degree of gene conservation in the datasets are represented by the same module colours. B. Visualizing the gene network heatmap plot. The network heatmap depicts the topological overlap matrix between among all genes. Yellow color represents low

overlap and darker red color represents higher overlap. Genotype maps and module assignments are also shown on the left and top.

Figure 2

Module-trait relationships. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and p-value. The table is color-coded by correlation according to the color legend.

Figure 3

Module membership and gene significance. A-D. Scatterplot of Gene Significance (GS) for stroke vs. module membership (MM) in the lightgreen, pink, yellow and midnightblue module

Figure 4

Validation for Hub genes regulatory network. Eight hub genes in the pink and yellow modules were colored with lightcyan, and the target genes were colored with magenta.